

# The fundamental diagram of urbanization

Giulia Carra

*Institut de Physique Théorique, CEA, CNRS-URA 2306, F-91191, Gif-sur-Yvette, France*

Marc Barthélemy\*

*Institut de Physique Théorique, CEA, CNRS-URA 2306, F-91191, Gif-sur-Yvette, France and  
Centre d'Analyse et de Mathématique Sociales, (CNRS/EHESS) 190-198,  
avenue de France, 75244 Paris Cedex 13, France*

The recent availability of geolocalized historical data allows to address quantitatively spatial features of the time evolution of urban areas. Here, we discuss how the number of buildings evolves with population and we show on different datasets (Chicago, 1930 – 2010; London, 1900 – 2015; New York City, 1790 – 2013; Paris, 1861 – 2011) that this curve evolves in a ‘universal’ way with three distinct phases. After an initial pre-urbanization phase, the first phase is a rapid growth of the number of buildings versus population. In a second regime, where residences are converted into offices and stores, the population decreases while the number of buildings stays approximatively constant. In another subsequent – modern – phase, the number of buildings and the population grow again and correspond to a re-densification of cities. We propose a simple model based on these simple mechanisms to explain the first two regimes and show that it is in excellent agreement with empirical observations. These results bring evidences for the possibility of constructing a simple model that could serve as a tool for understanding quantitatively urbanization and the future evolution of cities.

Keywords: Statistical Physics — Urbanization — City growth

## INTRODUCTION

Urbanization, measured by the fraction of people living in urban areas, gradually increased in modern countries with a quick growth since the middle of the 19<sup>th</sup> century until reaching values around 80% in most european countries [1]. This process depends in general on several economic variables and is connected to transportation technologies [2, 3]. Although urban development and the distribution of residential activity in urban areas are long-standing problems tackled by economists and geographers [4–12] a quantitative understanding of the different processes characterizing this phenomenon is still lacking. Among the first empirical analysis on population density, Meuriot [6] provided a large number of density maps of European cities during the nineteenth century, and Clark [13] proposed the first quantitative analysis of empirical data. In [14] Anas presents an economic model for the dynamics of urban residential growth, and in regional models describe the population dynamics of systems divided into zones characterised by a set of socio-economic indicators and that exchange with one another population, goods, capital, etc. according to some optimization rule. In this framework, the authors of [15] proposed a dynamical central place model highlighting the importance of both determinism and fluctuations in the evolution of urban systems. In [16], the author reviews different approaches used to model population dynamics in cities and in particular the ecological approach, where ideas from mathematical ecology models are used to study urban systems. An example is given by [17] where phase portraits of differential equations bring qualitative insights about urban systems be-

havior. Other important theoretical approaches comprise the classical Alonso-Muth-Mills model [2] developed in urban economics, and also numerical simulations based on cellular automata [18].

For most of these studies, numerical models usually require a large number of parameters that makes it difficult to test their validity and to identify the main mechanisms governing the urbanization process. On the other hand, theoretical approaches propose in general a large set of coupled equations that are difficult to handle and amenable to quantitative predictions that can be tested against data. In addition, even if a qualitative understanding is brought by these theoretical models, empirical tests are often lacking.

The recent availability of geolocalized, historical data (such as in [19] for example) from world cities [20] has the potential to change this state and allows to revisit with a fresh eye these long-standing problems. Many cities created open-data websites [21] and the city of New York (US) played an important role with the release of the PLUTO dataset (short for Property Land Use Tax lot Output), where tax lot records contain a lot of information about the urbanization process. For example, in addition to the location, property value, square footage etc, this dataset gives access to the construction date for each building. This type of geolocalized data at a very small spatial scale allows to monitor the urbanization process in time and at a very good spatial resolution.

These datasets allow in particular to produce ‘age maps’ where the construction date of buildings is displayed on a map (see Figure 1 for the example of the Bronx borough in New York City). Many age building maps are now available: Chicago [22]. New York

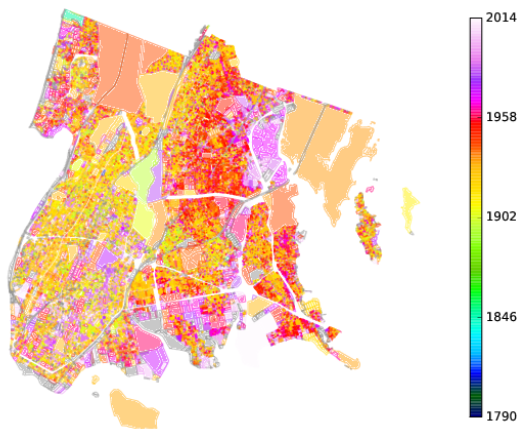


FIG. 1: Map of buildings construction date for the case of the Bronx (New York City, US). Most of the buildings were constructed during the beginning of the 20<sup>th</sup> century, followed by the construction in some localized areas of buildings in the second half of the 20<sup>th</sup> century. (See Material and Methods for details on the dataset).

City (US) [23], Ljubljana (Slovenia) [24], Reykjavik (Iceland) [25], etc. In addition to be visually attractive [26, 27], these maps together with new mapping tools (such as the urban layers proposed in [27]) provide qualitative insights into the history of specific buildings and also into the evolution of entire neighborhoods. Palmer [28] studied the evolution of the city of Portland (Oregon, US) from 1851 and observed that only 942 buildings are still left from the end of the 19th century, while 75,434 buildings were built at the end of the 20th century and are still standing, followed by a steady decline of new buildings construction since 2005. Inspired by Palmers map, Plahuta [29] constructed a map of building ages in his home town of Ljubljana, Slovenia, and proposed a video showing the growth of this city from 1500 until now [30]. Plahuta observed that the number of new buildings constructed each year displays huge spikes that signalled important events: an important spike occurred a few years after a major earthquake hit the area in 1899 when people were able to rebuild and other periods of rebuilding occurred after the two world wars. In the case of Los Angeles (US), the ‘Built:LA project’ shows the ages of almost every building in the city and allows to reveal the city growth over time [31].

These different datasets allow thus to monitor at a very small spatial resolution the urbanization process. In particular, for a given district or zone, we can ask quantitative questions about the evolution of the population and of the number of buildings. Surprisingly enough, such a dual information is difficult to find and – up to our knowledge – was not studied at the quantitative level. Here, we use data for different cities (Chicago, 1930 – 2010; London, 1900 – 2015; New York City, 1790 – 2013; Paris, 1861 – 2011) in order to answer questions about these

fundamental quantities. In particular, we will show that the number of buildings versus the population follows the same unique pattern for all cities studied here. We then propose an explanation for the existence of such a pattern and provide a theoretical model and empirical evidences supporting it.

## EMPIRICAL RESULTS

We investigate the urban growth of four different cities: Chicago (US), London (UK), New York (US), and Paris (France). An important discussion concerns the choice of the scale at which we study the urbanization process. We have to analyze the urbanization process at a spatial scale that is large enough in order to obtain statistical regularities, but not too large as different zones may evolve differently. Indeed geographers observed that the population density is not homogeneous and decreases in general with the distance to the center [32, 33]. Also, during the evolution of most cities, they tend to spread out with the density decreasing in central districts and increasing in the outer ones [13]. We thus choose to focus on the evolution of administrative districts of each city as at this level we can get insight about the growth process and exclude longer term processes. More precisely, we considered the 5 boroughs of New York, the 9 sides of Chicago, the 20 arrondissements of Paris and the 33 London districts. Also, in this way we do not have to tackle the difficult problem of city definition and its impact on various measures [34] and focus on the urbanization process of a given zone with fixed surface area.

The datasets for these cities come from different sources (see Materials and Methods) and cover different time periods. 1930 – 2010 for Chicago, 1900 – 2015 for London, 1790 – 2013 for New York, and 1861 – 2011 for Paris. An important limitation that guided us for choosing these cities is the simultaneous availability of building age and historical data for district population. In the following, in order to provide an historical context, we will first measure the evolution of the population density and then analyze the evolution of the number of buildings in a given district and its population.

### Population density growth

In order to get a first understanding of the urban growth behavior of these different cities, we begin with an empirical analysis of the evolution of urban density on a fixed geographical area. In Fig. 2 we show the average population density for the four cities studied here. This plot reveals that these different cities follow similar dynamics. After a positive growth and a population increase that accelerates around 1900, we observe a density peak. After this peak, the density decreases (even

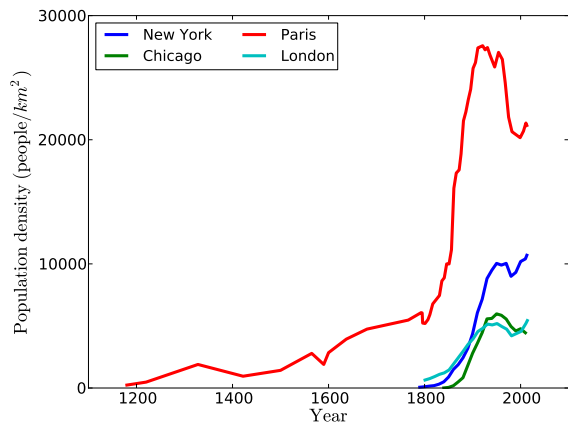


FIG. 2: **Population density versus time.** The average population density versus time for the four cities studied in the paper. All these cities display a density peak in the first half of the 20<sup>th</sup> century (see Material and Methods for details on datasets).

sharply in the case of NYC) or stays roughly constant. In the last years, New York City, Paris and London display a re-densification period. This first figure highlights the existence of a seemingly ‘universal’ pattern governing the urbanization process. These large cities are divided in districts that usually display different properties and in Fig. 3 we plot the time evolution of some district densities. We indeed observe that they display different behaviors. In the case of London (Fig. 3, top panels),

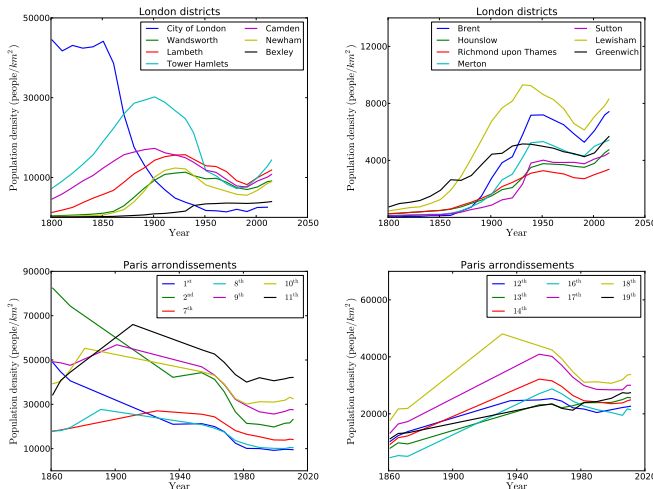


FIG. 3: **Population density versus time.** Local population densities for a selection of London districts (top), and a selection of Paris arrondissements (bottom). For the sake of clarity we did not plot all the districts studied and additional results can be found in the SI.

we note that the district City of London reached a density peak before 1800 while other districts (for example Lewisham, Brent and Newham) display all the different

phases of urbanization described above. For Chicago (see SI) and Paris (Fig. 3, bottom panels) for Paris), the different districts are not all synchronized and display simultaneously different urbanization phases. The central districts of Paris (the 1<sup>st</sup> and the 4<sup>th</sup> for example) typically reached their density peak before 1860, while less central districts (11<sup>th</sup> to 20<sup>th</sup>) reached their density peak in the first half of the 20<sup>th</sup> century, consistently with the idea of a centrifugal urbanization process.

For the five boroughs of New York (see SI), we observe that Manhattan (MN), the Bronx (BX) and Brooklyn (BK) already passed through the different phases of urbanisation, and are now in a re-densification period. In contrast, Staten Island (SI) and Queens (QN) are still in the urbanisation period characterized by a positive population growth rate and didn’t reach yet the density peak.

These preliminary results highlight the importance of spatial delimitations when studying a city. The dynamics of different districts might be the same but are not necessary simultaneous. For this reason, we will not consider in the following cities as a whole, but rather follow the evolution of various quantities for each district which display a better level of homogeneity.

### Number of building vs. population

A crucial quantity that measures the level of urbanization is the number of buildings in a given area. For each district, we can then study the relation between the number of buildings  $N_b$  and the population  $P$  of different districts (Fig. 4). We observe an apparent diversity of behaviors but, as we will see in the following, they can all be described within a single ‘fundamental diagram’. In Fig. 4 top-left we show the result for the five boroughs of New York City. We observe that SI and QN (dashed lines) are in a growing phase characterized by a positive value of  $dN_b/dP$ , while Manhattan, Brooklyn and Bronx (plotted in continuous line) reached other dynamical regimes. In Fig. 4 top-center-left we plot the nine sides of Chicago, and we observe a clear growth phase followed by a ‘saturation’ (corresponding to the density peak) for the Far North, Northwest, Southwest, Far Southeast and Far Southwest sides (plotted in continuous line). In contrast, the others sides (Central, North, West and South), in dotted line, seem to have reached a saturation before 1930. Indeed, the dotted lines (that have to be read chronologically from the right to the left) do not display the growth regime, suggesting that it stopped before 1930, year of the earliest available data. In Fig. 4 top-centre-right, we represent the evolution for some Paris arrondissements. We observe the growth regime followed by a saturation for the 10<sup>th</sup>, 12<sup>th</sup>, 16<sup>th</sup> and 18<sup>th</sup> arrondissement (in continuous line), the 13<sup>th</sup> seem not having reach a saturation yet, while the others have reached saturation before 1861. In the top-

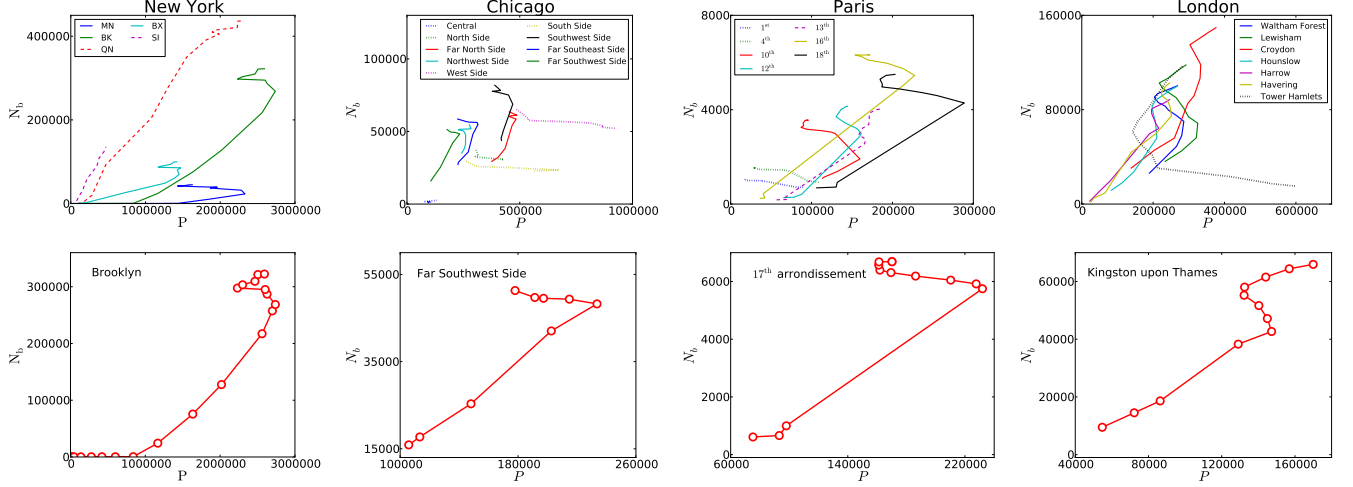


FIG. 4: **Number of buildings versus population.** We represent with continuous lines the districts that have reached their density peak, with dashed lines for districts that are still in the growing phase. We use dotted line for the districts that reached the density peak before the first year available in the dataset. (Top panels) Results for districts in the cities studied here. (Bottom) We show examples illustrating the ‘universal’ diagram for the four different cities.

right plot of Fig. 4 about London districts, we observe that all districts displayed here reached a saturation, but that the district Tower Hamlets (dotted line) reached it before 1900, year of the first available data.

These various plots show that for different districts we have essentially the same trajectory in the plane  $(P, N_b)$ . We show illustrative examples for various cities in Fig. 4(bottom). The evolution of the districts of these different cities can thus be represented by a typical path shown in Fig. 5. This ‘fundamental diagram’ is characterized by a first phase of rapid growth of the number of buildings versus population. In a second regime, the population decreases while the number of buildings stays roughly constant. In a last – modern – phase, the number of buildings and population grow again.

### THEORETICAL MODEL

The data studied in the previous section display a pattern that seems to encompass specific features of the different cities and we propose a theoretical model based on the following interpretation for these different regimes. The first regime corresponds to the urbanization where buildings are constructed on empty lots until the ‘saturation point’  $(P^*, N_b^*)$ , which signals the beginning of the second regime. In this second regime, residences are converted into stores and offices, and the population naturally decreases while the number of buildings stays approximately constant. In the last regime, both the number of buildings and the population grow again, corresponding to the ‘re-densification’ of cities. In order to provide quantitative evidences for the first two phases of this process, we propose a simple model based on this

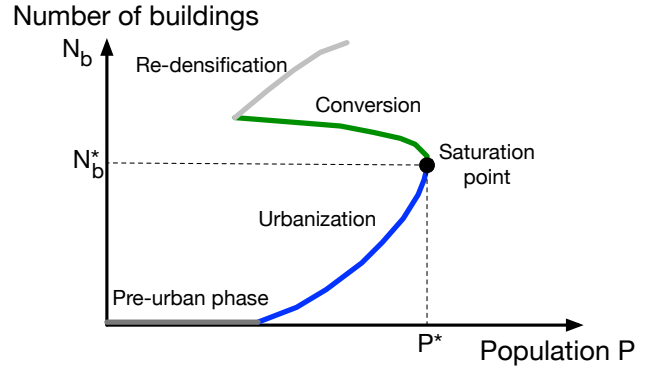


FIG. 5: **Schematic representation of the fundamental curve.** We represent here the typical district growth curve characterized by three main phases: after a pre-urbanization period, there is first an urbanization phase with a positive growth rate  $dN_b/dP$  that stops at the ‘saturation point’  $(P^*, N_b^*)$ . A second ‘conversion’ phase follows, during which the population decreases. Finally, we observe a last re-densification phase where both the population and the number of buildings increase.

interpretation, and that we will test against data. We model the evolution of a given zone of surface area  $A$  by a two-dimensional square grid where each cell of surface  $a_\ell$  represents an empty, constructible lot. The maximum number of lots is then given by  $N_{max} = A/a_\ell$ . Each cell can be empty or occupied (a building has already been built) and each building on a lot  $i$  is characterized by its number of residential floors  $h_r(i)$ , commercial floors  $h_c(i)$  (the total number of floors is  $h(i) = h_r(i) + h_c(i)$ ). At each time step  $t \rightarrow t + \Delta t$  (in the following we count the

time  $t$  in units of  $\Delta t$ ), we pick at random a cell  $i$  and if it is empty we update it with

$$\begin{cases} P \rightarrow P + \Delta P, \\ N_b \rightarrow N_b + 1, \\ h(i) = h_r(i) = 1, \\ h_c(i) = 0, \end{cases} \quad (1)$$

where  $\Delta P$  is the number of people per residential floor. If a building is already present on the chosen cell, we add an extra residential floor with probability  $p_h$  or convert a floor into a non-residential one (such as offices or stores) with probability  $p_c$ :

$$\begin{cases} h_r(i) \rightarrow h_r(i) + 1 \\ h_c(i) \rightarrow h_c(i) \\ P \rightarrow P + \Delta P \end{cases} \quad \text{with prob. } p_h \quad (2)$$

$$\begin{cases} h_r(i) \rightarrow h_r(i) - 1 \\ h_c(i) \rightarrow h_c(i) + 1 \\ P \rightarrow P - \Delta P \end{cases} \quad \text{with prob. } p_c \quad (3)$$

Finally, nothing happens with probability  $1 - p_h - p_c$ . Each district is thus characterized by the parameters  $\Delta P$ ,  $p_c$  and  $p_h$ . The mean-field equations describing the evolution of  $H_r = \sum_i h_r(i)$  (the total number of residential floors in the district), the total number of buildings  $N_b$  and the total population  $P$  in the district are

$$\frac{dH_r}{dt} = \frac{N_b}{N_{max}}(p_h - p_c) + \left(1 - \frac{N_b}{N_{max}}\right), \quad (4)$$

$$\frac{dN_b}{dt} = 1 - \frac{N_b}{N_{max}}, \quad (5)$$

$$\frac{dP}{dt} = \Delta P \frac{dH_r}{dt}. \quad (6)$$

Solving Eq. (5) and Eq. (6) leads to

$$N_b(t) = N_{max} \left(1 - e^{-t/N_{max}}\right), \quad (7)$$

$$P(t) = \Delta P \left[(p_h - p_c)t + N_{max}(1 + p_c - p_h)(1 - e^{-t/N_{max}})\right] \quad (8)$$

Eq. (6) implies that the population is an increasing function of the number of building up to a saturation value  $N_b^*$  corresponding to the population  $P^*$ , after which the population decreases (ie. above which  $dP/dt$  becomes negative). After simple calculations (see SI for details), we obtain

$$N_b^* = \frac{N_{max}}{1 + p_c - p_h}, \quad (9)$$

$$\frac{P^*}{\Delta P N_{max}} = (p_h - p_c) \log \left( \frac{1 + p_c - p_h}{p_c - p_h} \right) + 1. \quad (10)$$

The saturation happens only if  $N_b^* < N_{max}$  and thus if  $p_c > p_h$  which expresses the fact that the conversion rate

should be large enough in order to observe a saturation point (if the conversion rate is too small, the first phase of growth will continue indefinitely). Defining the normalized variables  $N_b^{*'} = N_b^*/N_{max}$  and  $P^{*'} = P^*/N_{max}$ , we can rewrite the above equation as

$$P^{*'} = \Delta P \left[1 + (1/N_b^{*'} - 1) \log(1 - N_b^{*'})\right]. \quad (11)$$

This relation allows us to determine the average number of people per building floor  $\Delta P$  for each district (see the SI for a discussion about this parameter). The theoretical results given by Eq. (7) and Eq. (8) imply a scaling that can be checked empirically. Indeed, if we make the following change of variables

$$\begin{aligned} X(t) &= \frac{N_b(t)}{N_{max}}, \\ Z(t) &= \frac{\frac{P(t)}{\Delta P N_{max}} - \frac{N_b(t)}{N_b^*}}{\frac{1}{N_b^{*'}} - 1}, \end{aligned} \quad (12)$$

then the curves for the different districts at different times should all collapse on the same curve given by

$$Z = \log(1 - X). \quad (13)$$

In order to test this model, we focus on districts that have already reached saturations (the others are still in the first growth phase). From the data we know the area  $A$  of each district and the average building footprint surface  $a_l$  of each district. This allows us to compute the maximum number of buildings  $N_{max} = A/a_l$  of the district. Moreover, the empirical curves allow us to determine the saturation values  $(P^*, N_b^*)$ , corresponding to the value of the population and the number of buildings after which the density growth rate becomes negative, from which we compute  $(P^{*'}, N_b^{*'})$ . At this point, we thus have estimated from empirical data all the parameters that characterize a district, without performing any fit. We can now test the scaling Eq. (13) predicted by the model. As explained above, the curves obtained for different districts should all collapse on the theoretical one. In Fig. 6 we plot the theoretical prediction (red line) and the values for the different districts (represented by different symbols and different colors for the different districts). An excellent collapse is observed, supporting the validity of the model.

## DISCUSSION AND PERSPECTIVES

Theoretical urban models can be roughly divided in two categories. On one hand there are economics models characterized by complex mathematical equations rarely amenable to quantitative predictions that can be tested against data. On the other hand, there are computer simulations (such as agent-based or cellular automata) that

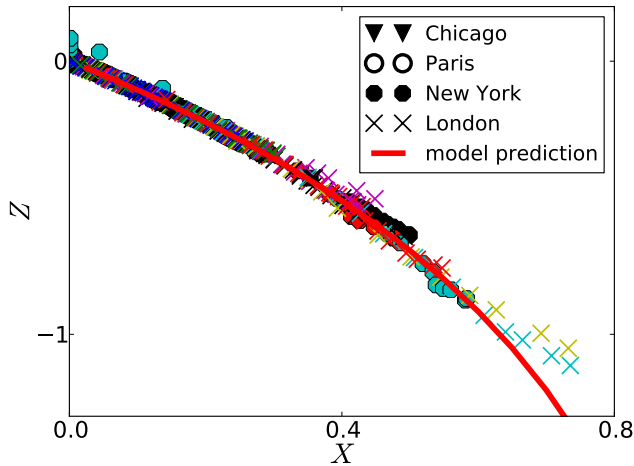


FIG. 6: **Collapse for the rescaled variable  $Z$  and  $X$ .** We plot the rescaled variables  $Z$  versus  $X$  (Eqs. (12)) for all the saturated districts of all cities. Each city is characterized by a different symbol and each district by a different color. The continuous red line is the theoretical prediction given by Eq. (13).

are characterized by a large number of parameters, preventing to understand the hierarchy of processes governing the phenomenon. In the approach presented here, we build a simple model with the smallest number of parameters and able to describe quantitatively the evolution of various quantities such as the number of buildings and the population for a given district. Our empirical analysis shows that there are essentially three different phases of the urbanization process: a growth phase where we observe an increase of both the number of buildings and the population; a second regime where the population decreases while the number of buildings stays roughly constant, and a last phase where both population and the number of buildings are increasing. The first two phases are well described by a simple model which integrate the crucial ingredient of converting residential space into commercial activities. We observe empirically the existence of a ‘re-densification’ phase where both population and the number of buildings increase after the conversion phase. This phase seems to happen simultaneously for the different districts in a city which suggests that it is an effect due to planning decisions and not resulting from self-organization. Modelling the appearance of this regime is thus at this point a challenge for future studies.

Beside showing that a minimal modeling for describing urbanization is possible despite the large variety of cities, we believe that this approach could constitute the basis for more elaborated models. These models could then be thoroughly tested against data, could describe the impact of various parameters and also help to understand some features of the possible future evolution of cities.

## MATERIALS AND METHODS

### Data description

#### *New York data*

We used data from the Primary Land Use Tax Lot Output (PLUTO) data file, developed by the New York City Department of City Plannings Information Technology Division (ITD)/Database and Application Development Section [35]. It contains extensive land use and geographic data at the tax lot level. PLUTO data files contain three basic types of data: tax lot characteristics, building characteristics and geographic/political/administrative districts. In particular for each building of the city we focused on the building’s borough, the building age and the surface of the lot. For each borough we compute the average building surface  $a_l$  (assumed to be given by the average building lot surface) over all the buildings in the borough and with known age (for New York city we have this information for 94% of buildings). New York data cover the period from 1790 to 2013. For the historical population data, we used different sources [36–39]

#### *Chicago data*

We used the Building Footprints dataset (deprecated August 2015) provided by the Data portal of the City of Chicago [40]. For each building we have the information on the geometrical shape from which we compute the building surface, the year built and the position. By using the shapefiles of the 77 Chicago communities [41], we can deduce the community (and thus the side) where the building is located in. For each side we compute the average building surface  $a_l$  and average this quantity over all the buildings with known year built, situated in the side. For Chicago the percentage of buildings with known built year is 54%. Population data from each community area comes from [42] and they cover the period from 1930 to 2010.

#### *Paris data*

We used the dataset ‘Emprise Batie Paris’ provided by the open data initiative of the ‘Atelier Parisien d’urbanisme (APUR)’ [43]. For each building we have the information on the geometrical shape, from which we compute the building surface, the year built and the arrondissement the building is situated in. For each arrondissement we compute the average building surface  $a_l$  averaging this quantity over all the buildings with known year built (i.e. the 57% of the buildings), situated in the arrondissement. Population data comes from [44] and



since the actual arrondissements were defined in 1859, population data at the level of the arrondissements covers the period from 1861 to 2011.

#### London data

We used the dataset ‘Dwelling Age Group Counts (LSOA)’ [45], which contain the residential dwelling ages, grouped into approximately 10-year age bins from pre-1900 to 2015 (the bin 1940 – 1944 is missing). The number of properties is given for each LSOA area and each age bin. From these data we deduced the number of buildings for each London district as function of the year. Data for the historical population of the London boroughs were obtained from ‘A Vision of Britain through time’ [46]. Finally we used OSOpenMapLocal [47] containing the geometrical shape of the buildings in London for computing the average footprint surface for each district. We note that in this last dataset some buildings are aggregated and rendered as homogenized zones. For this reason we computed the average building surface of each district by averaging over all the buildings belonging to the district having a footprint surface smaller than  $700m^2$ . In order to locate the district to which a building belongs to, we used the shapefile of London districts boundaries [48].

#### ACKNOWLEDGEMENTS

GC thanks the Complex Systems Institute in Paris (ISC-PIF) for hosting her during part of this work. MB thanks the program Paris 2030 for financial support.

---

\* Electronic address: marc.barthelemy@cea.fr

- [1] M. Antrop, *Landscape change and the urbanization process in Europe*, Landscape and urban planning 67.1 (2004): 9-26.
- [2] A. Anas, R. Arnott., K.A. Small, *Urban spatial structure*, Journal of economic literature 36.3 (1998): 1426-1464.
- [3] E.L. Glaeser and M.E. Kahn, *Sprawl and Urban Growth*, the National Bureau of Economic Research, Working Paper 9733 <http://www.nber.org/papers/w9733> (2003).
- [4] J.D. Herbert and B.H. Stevens, *A model for the distribution of residential activity in urban areas*. Journal of Regional Science, 2-2 (1960).
- [5] E.S. Mills, *An Aggregative Model of Resource Allocation in a Metropolitan Area*. The American Economic Review, 57-2 (1967), pp. 197-210.
- [6] Meuriot, M.P. *Des Agglomerations Urbaines Dans L’europe Contemporaine: Essai Sur Les Causes, Les Conditions, Les Consequences De Leur Developpement*. Paris: Belin, (1898).
- [7] Taylor, T.G. *Urban geography*. London: Methuen, (1949).
- [8] W.C. Wheaton, *Urban Residential Growth under Perfect Foresight* Journal of Urban Economics, 12 (1982), pp. 1–21.
- [9] M.J. Beckmann, *On the distribution of urban rent and residential density* Journal of Economic Theory, 1-1 (1969), pp. 60–67.
- [10] D. Harrison and J.F. Kain, *Cumulative Urban Growth and Urban Density Functions*. Journal of Urban Economics, 1 (1974), pp. 61–98.
- [11] Tobler, Waldo R. *A computer movie simulating urban growth in the Detroit region*. Economic geography 46.sup1 (1970): 234-240.
- [12] Makse H.A., Havlin S., Stanley H.E., *Modelling urban growth*. Nature, 337:1912 (1995), 779-782.
- [13] C. Clark, *Urban population densities*. Journal of the Royal Statistical Society, Series A (General), 114-4 (1951), pp. 490–496.
- [14] A. Anas, *Dynamics of Urban Residential Growth*, Journal of Urban Economics, 5 (1978), pp. 66–87.
- [15] P.M. Allen and M. Sanglier, *Urban evolution, self-organization, and decision making*, Environment and Planning A, 13 (1981), pp. 167–183.
- [16] I. Benenson, *Modeling population dynamics in the city: from a regional to a multi-agent approach*, Discrete dynamics in nature and society, 3 (1999), pp. 149–170.
- [17] D. Dendrinos and H. Mullally, *Evolutionary patterns of urban populations*, Geographical Analysis, 13 (1982), pp. 328–344.
- [18] D. O’Sullivan and P.M. Torrens, *Theory and Practical Issues on Cellular Automata* chapter *Cellular Models of Urban Systems*, Springer London (2001). url=”[http://dx.doi.org/10.1007/978-1-4471-0709-5\\_13](http://dx.doi.org/10.1007/978-1-4471-0709-5_13)”
- [19] J. Perret, M. Gribaudi, M. Barthelemy, *Roads and cities of 18th century France* Scientific data 2 (2015).
- [20] Angel, S., J. Parent, D.L. Civco and A.M. Blei. *The Atlas of Urban Expansion*. Lincoln Institute of Land Policy (2012).
- [21] Mason B. (2013) *Brilliant Maps Reveal Age of the Worlds Buildings*. Available at <http://www.wired.com/2013/10/building-ages-map-gallery/>. Accessed 31/08/2016.
- [22] Jacobsen S. (2013) *Chicago Building Age Map*. Available at [http://transitized.com/chibld/index.html?utm\\_source=//Transitized&utm\\_medium=post&utm\\_campaign=New%20CBAM%20Post](http://transitized.com/chibld/index.html?utm_source=//Transitized&utm_medium=post&utm_campaign=New%20CBAM%20Post) (Accessed 31/08/2016).
- [23] Brandon L. *Building Age NYC*. Available at <http://pureinformation.net/building-age-nyc/#12/40.7392/-73.9651>. Accessed 31/08/2016.
- [24] Available at <http://www.virostatiq.com/data/ljubljana-building-ages/>. Accessed 31/08/2016.
- [25] Riggott M. *The Age of Greater Reykjavík*. Available at <http://tiles.flother.is/2013/reykjavik-age/>. Accessed 31/08/2016.
- [26] Palmer J. (2014) *Portland Oregon: The Age of a City*. Available at <http://labratrevenge.com/pdx/#12/45.4483/-122.7139>. Accessed 31/08/2016.
- [27] Morphocode *Urban Layers*. Available at <http://io.morphocode.com/urban-layers/>. Accessed 29/08/2016.
- [28] Palmer J. (2013) *Portland Oregon: The Age of a City*. Available at <http://dealloc.me/2013/06/30/the-making->

- of-pdx/. Accessed 31/08/2016.
- [29] Plahuta M. (2013) *Building ages in Ljubljana, Slovenia*. Available at <http://virostatiq.com/structure-ages-in-ljubljana-slovenia/>. Accessed 29/08/2016.
- [30] O'Hara M. *City of Ljubljana - growth between years 1500 - 2013*. Available at <https://vimeo.com/72249300>. Accessed 29/08/2016.
- [31] Ureta O. *built: LA Building Age // 1890-2008*. Available at [http://cityhubla.github.io/LA\\_Building\\_Age/#12/34.0267/-118.2621](http://cityhubla.github.io/LA_Building_Age/#12/34.0267/-118.2621). Accessed 29/08/2016.
- [32] Bertaud A. and Malpezzi S. (2003). *The spatial distribution of population in 48 world cities: Implications for economies in transition*. Report, World Bank.
- [33] Gu erois M., Pumain D. (2008). *Built-up encroachment and the urban field: a comparison of forty european cities*. Environment and Planning A 40: 2186-2203.
- [34] Arcaute, E., Hatna, E., Ferguson, P., Youn, H., Johansson, A., Batty, M. (2015). Constructing cities, deconstructing scaling laws. *Journal of The Royal Society Interface*, 12(102), 20140745.
- [35] *Primary Land Use Tax Lot Output (PLUTO) dataset*, developed by the New York City Department of City Planning's Information Technology Division (ITD)/Database and Application Development Section. Retrieved from <http://www1.nyc.gov/site/planning/data-maps/open-data.page>.
- [36] *Table PL-P1 NYC: Total Population New York City and Boroughs, 2000 and 2010 (PDF)*. nyc.gov. Retrieved 16 May 2016.
- [37] U.S. Census Bureau, Population Division, Table 5. Annual Estimates of the Resident Population for Minor Civil Divisions in New York, Listed Alphabetically Within County: April 1, 2000 to July 1, 2009 (SUB-EST2009-05-36) and Table 1. Annual Estimates of the Resident Population for Incorporated Places Over 100,000, Ranked by July 1, 2009 Population: April 1, 2000 to July 1, 2009 (SUB-EST2009-01), Release Date: June 2010, retrieved on July 31, 2010.
- [38] Forstall, Richard L., Population of States and Counties of the United States: 1790 to 1990, U.S. Bureau of the Census, Washington, DC, 1996 ISBN 0-934213-48-8, (Part III, Kentucky to Oklahoma) retrieved April 3, 2011.
- [39] "Population", article by Jane Allen with tables by Nathan Kantrowitz in *The Encyclopedia of New York City*, edited by Kenneth T. Jackson, New-York Historical Society and Yale University Press, 1995, pages 910-914, ISBN 0-300-05536-6.
- [40] City of Chicago Data Portal. *Building Footprints (deprecated August 2015)*. Available at <https://data.cityofchicago.org/Buildings/Building-Footprints-deprecated-August-2015-/qv97-3bvb/data>. Accessed 31/08/2016
- [41] City of Chicago Data Portal. *Boundaries - Community Areas (current)*. Available at <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6/data>. Accessed 31/08/2016.
- [42] <http://www.robparal.com/downloads/ACS0509/HistoricalData/Chicago>.
- [43] Plateforme open data de l'atelier parisien d'urbanisme. *Emprise batie Paris*. Available at [http://cassini.apur.opendata.arcgis.com/datasets/002f14-c0cf28435296a341d9921adf99\\_0](http://cassini.apur.opendata.arcgis.com/datasets/002f14-c0cf28435296a341d9921adf99_0). Accessed 31/08/2016
- [44] Data from Wikipedia. Available at [https://fr.wikipedia.org/wiki/1er\\_arrondissement\\_de\\_Paris](https://fr.wikipedia.org/wiki/1er_arrondissement_de_Paris). Accessed 31/08/2016.
- [45] Consumer Data Research Centre, *Dwelling Age Group Counts (LSOA)*, retrieved from <https://data.cdrc.ac.uk/dataset/house-ages-and-prices/resource/4f1956b2-3128-4297-ba97-059e1fbc1fcc>. Accessed 31/08/2016.
- [46] A vision of Britain through time. Available at <http://www.visionofbritain.org.uk/>. Accessed 31/08/2016.
- [47] Ordnance Survey, *OS Open Map - Local*, retrieved from <https://www.ordnancesurvey.co.uk/business-and-government/products/os-open-map-local.html>. Accessed 31/08/2016.
- [48] London DataStore. <http://data.london.gov.uk/dataset/-statistical-gis-boundary-files-london>



## SUPPLEMENTARY INFORMATION

### The model: calculations

We analyze here Eq. [6]. In particular, we show that there is in this model a critical value of  $N_b$  above which the population decreases (ie. above which  $dP/dt$  becomes negative). We thus solve:

$$\frac{dP}{dt} \geq 0$$

$$\Delta P \frac{N_b}{N_{max}} (p_h - p_c) + \Delta P (1 - \frac{N_b}{N_{max}}) \geq 0 ,$$

and obtain the following condition

$$N_b \leq \frac{N_{max}}{1 + p_c - p_h} \quad (14)$$

which then implies that

$$N_b^* = \frac{N_{max}}{1 + p_c - p_h} . \quad (15)$$

We thus observe a saturation effect only if  $N_b^* < N_{max}$  and thus if  $p_c > p_h$ . We can then compute the time  $t^*$  for which saturation happens: knowing that

$$N_b^* = N_{max} (1 - e^{-t^*/N_{max}}) \quad (16)$$

we get

$$t^* = N_{max} \log \left( \frac{1 + p_c - p_h}{p_c - p_h} \right) . \quad (17)$$

Using Eq. [8], we then obtain

$$P^* = \Delta P (p_h - p_c) N_{max} \log \left( \frac{1 + p_c - p_h}{p_c - p_h} \right) + \Delta P N_{max} . \quad (18)$$

### Additional measures

Analysing the relation between the number of buildings and the population during a city district growth, we observed the emergence of a 'universal' pattern characterized by three regimes: urbanization, conversion and densification. We present here additional measures and results for the different cities considered in this study.

#### *Chicago sides*

Concerning Chicago, 5 sides saturated, three of them in 1970 and two of them in 1960. Just one of them began a re-densification process in 1980. In the converting period we have a variation of the population equals to  $-16992$  corresponding to a decrease of 6.5% of the population.

#### *Paris arrondissements*

The first four arrondissements seem to have saturated before the first data available. The 13<sup>th</sup> seems not yet saturated and the others with the exception of the 6<sup>th</sup> present all three phases even if often the re-densification one is quite recent.

The average value of the saturation year is

$$Year_s = 1932 \pm 28 ,$$

the average value of the re-densification date is

$$Year_d = 1993 \pm 10 ,$$

the average period of conversation in years is

$$\Delta t_c = 60 \pm 30 ,$$

the average period of conversation in population is

$$\Delta P_c = -58940 \pm 25434 ,$$

that corresponds to an average decrease of

$$\frac{\Delta P_c}{P^*} = -0.33 \pm 0.16\% .$$

The behaviors seem thus quite various.

#### *London districts*

Eight of the 33 London districts seem saturated before the first data available. The others show all the three regimes.

We have

$$Year_s = 1949 \pm 15 ,$$

$$Year_d = 1992 \pm 2.43 ,$$

$$\Delta t_c = 43 \pm 15 ,$$

$$\Delta P_c = -63054 \pm 62908 ,$$

$$\frac{\Delta P_c}{P^*} = -0.2 \pm 0.15\% .$$

In particular one remarks that 23 of the 25 districts have  $Year_d = 1992$ .

#### *New York boroughs*

In New York city only three of the five boroughs reached saturation and all of these present a densification regime. We have

$$Year_s(MN) = 1910 \quad Year_s(BK) = 1950 \quad Year_s(BX) = 1970 ,$$

all the boroughs began de re-densification phase in 1980.

Measuring the parameter  $\Delta P$

The parameter  $\Delta P$  introduced in the model, defined as the number of people per residential floor has been estimated as explained in the main text from the equation

$$\frac{P^*}{N_{max}} = \Delta P [1 + (N_{max}/N_b^* - 1) \log (1 - N_b^*/N_{max})] = \Delta P / f(N_b^*, N_{max}). \tag{19}$$

In the tables below we reported for each district the estimation we obtained.

*New York*

borough	$\Delta P$
MN	184.84
BK	16.73
BX	29.36

*Paris*

arrondissement	$\Delta P$
7	104
8	146
9	131
10	159
11	149
12	103
13	116
14	85
15	98
16	76
17	75
18	128
19	118
20	127

*Chicago*

side	$\Delta P$
Far North Side	16.1
Far Southeast Side	11.2
Northwest Side	10.2
Far Southwest Side	9.2
Southwest Side	13

*London*

district	$\Delta P$
Lambeth	10.7
Greenwich	15
Sutton	6.6
Lewisham	8
Barnet	6.4
Hammersmith and Fulham	6.6
Barking and Dagenham	6.6
Enfield	6.7
Croydon	5.2
Merton	6.4
Haringey	6.9
Harrow	6.3
Hounslow	7.1
Kingston upon Thames	6.5
Havering	6.5
Waltham Forest	7
Hillingdon	5.8
Bexley	5.4
Ealing	6.1
Bromley	5.6
Redbridge	6.4
Newham	14.9
Wandsworth	8.8
Richmond upon Thames	6.3
Brent	6.9

**Additional empirical results**

For the sake of completeness, we present here additional results for the cities studies here.

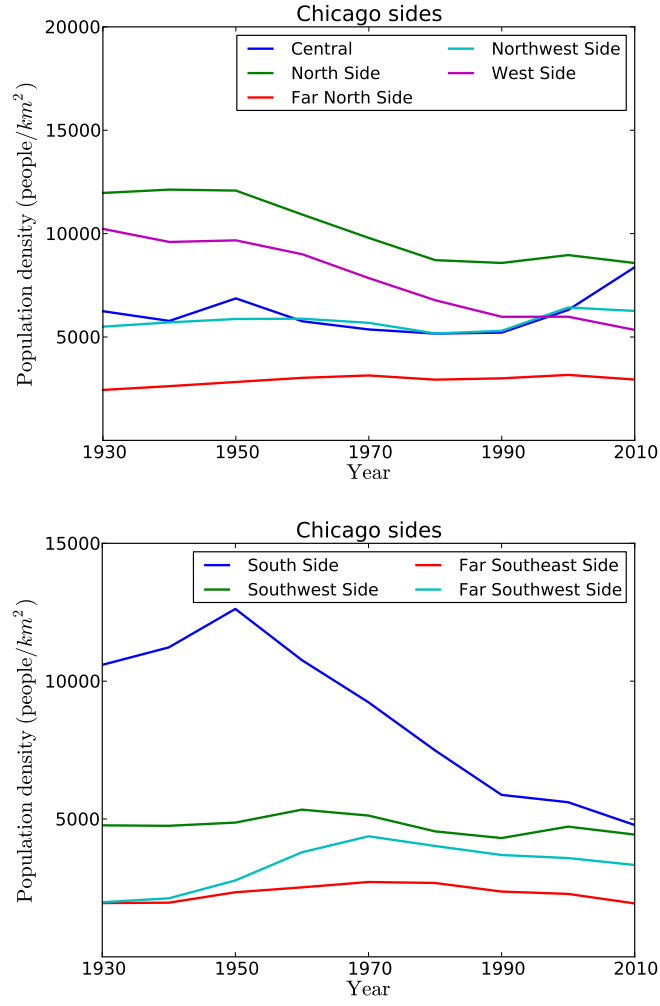


FIG. 7: Chicago districts: population density VS year

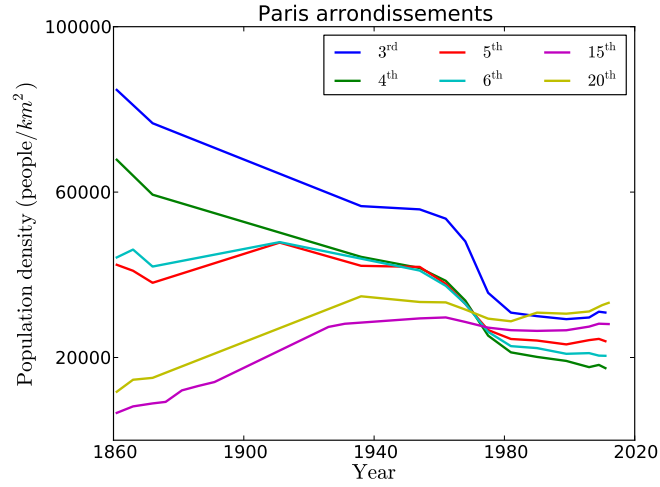


FIG. 8: Paris arrondissements: population density VS year

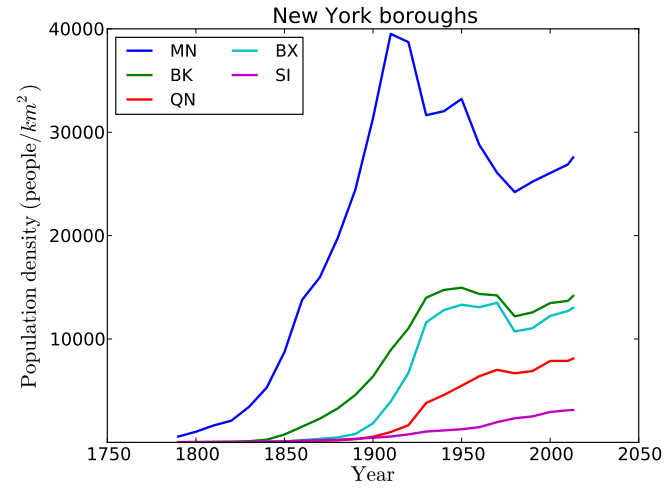


FIG. 9: New York boroughs: population density VS year



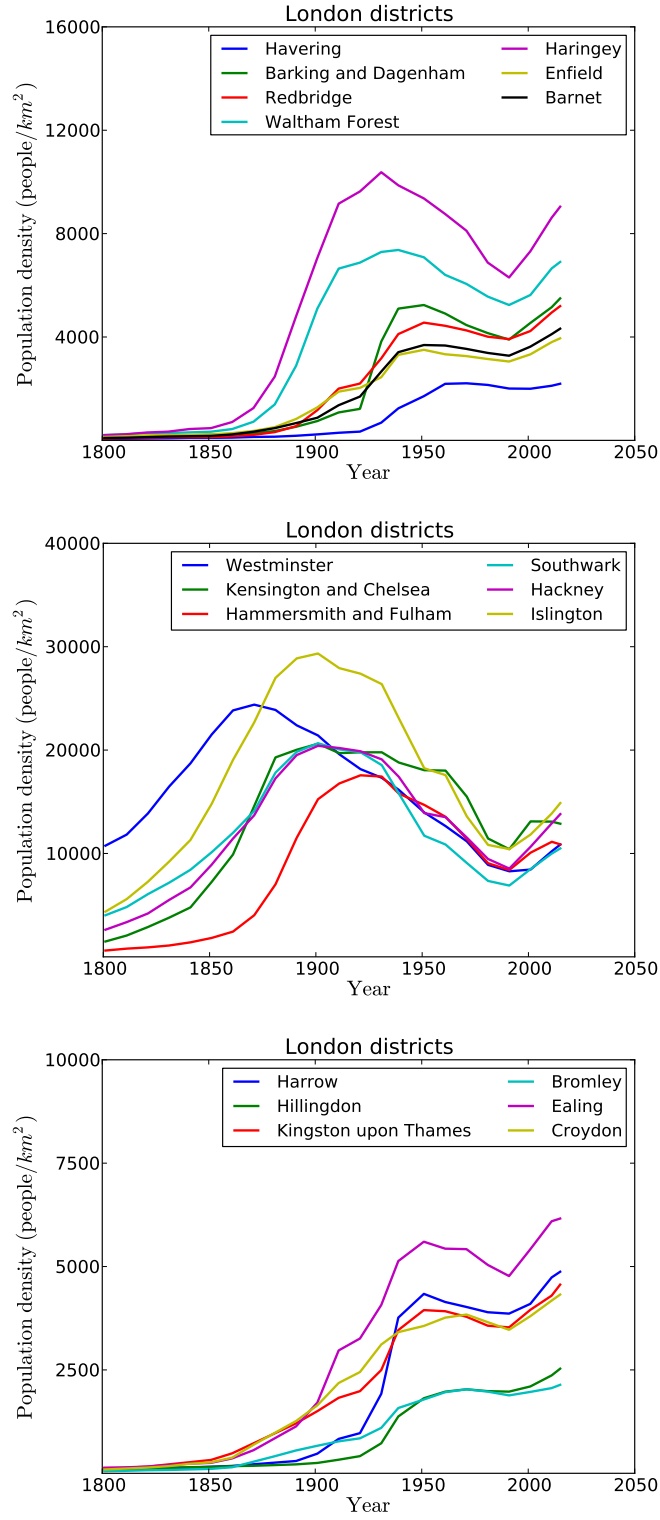


FIG. 10: London districts: population density VS year

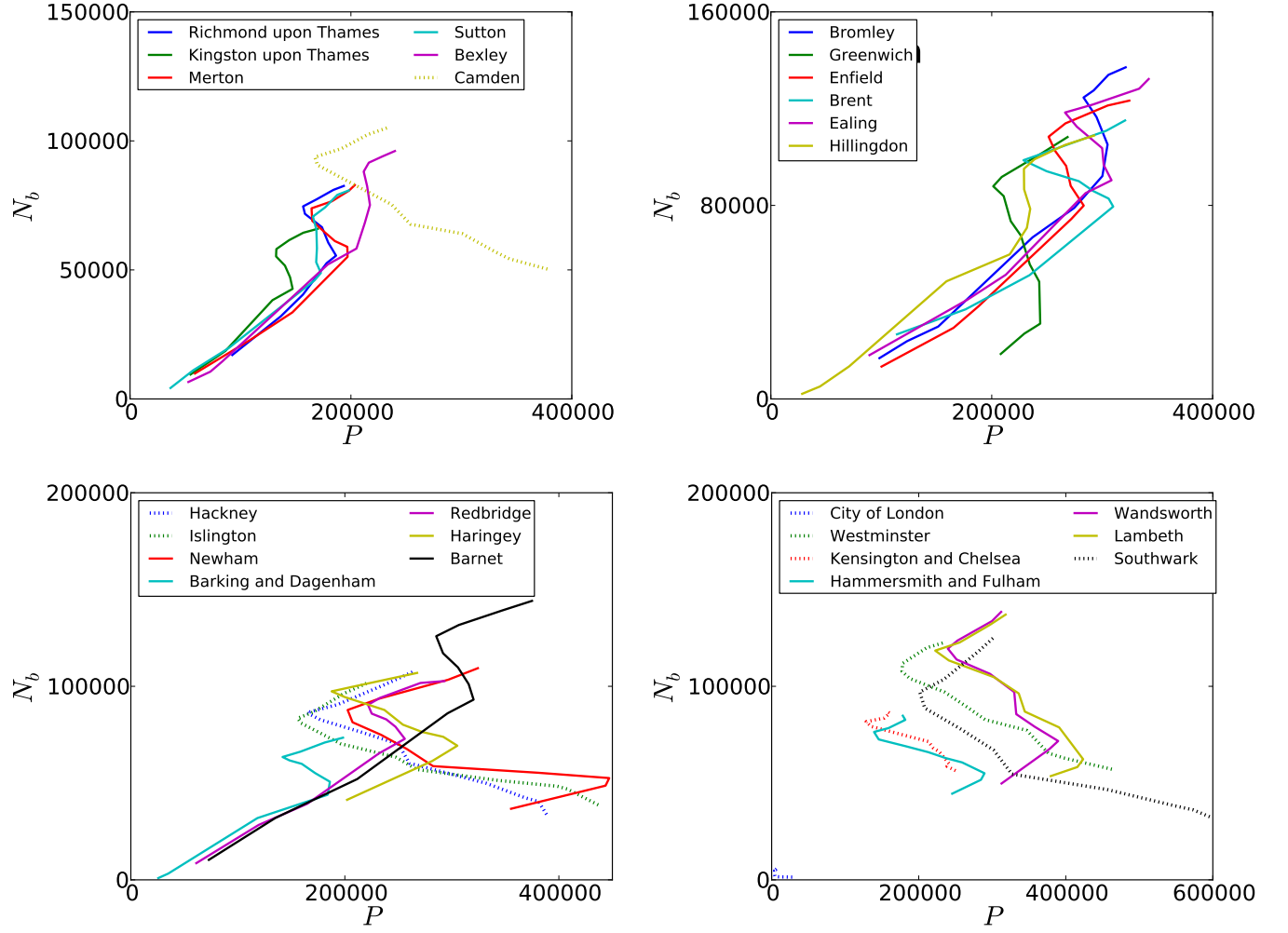


FIG. 11: **London districts: number of buildings VS population.** In continuous line we have districts that reached the saturation point. In dashed line we have districts that are still in the growing phase and in dotted line the ones that reached the saturation points before the year of the first available data.

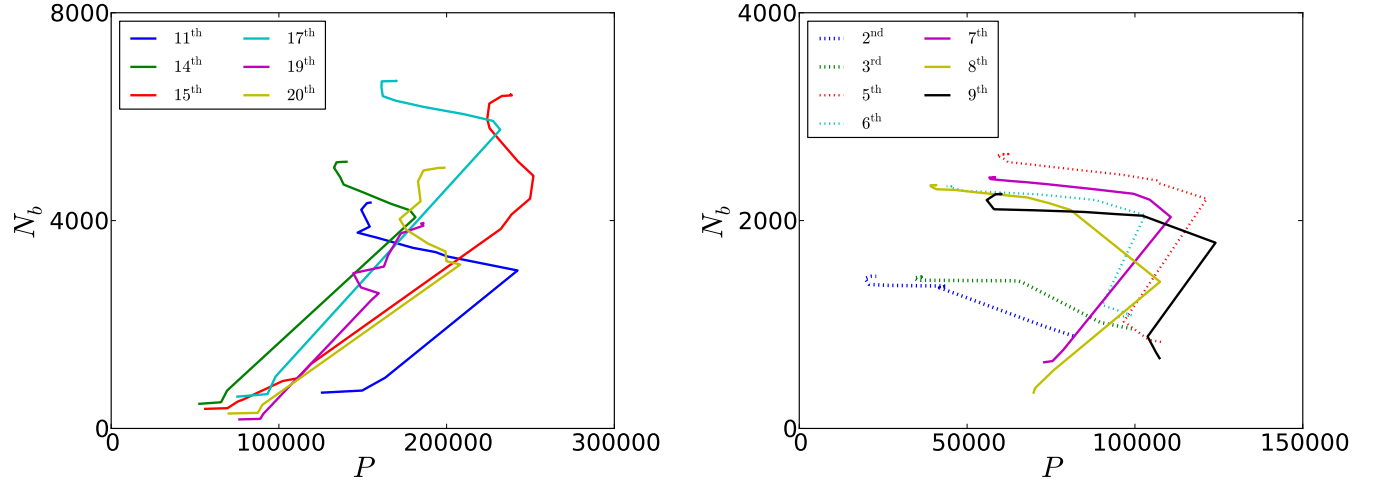


FIG. 12: **Paris arrondissements: number of buildings VS population.** In continuous line we have districts that reached the saturation point. In dashed line we have districts that are still in the growing phase and in dotted line the ones that reached the saturation points before the year of the first available data.